

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平5-181912

(43) 公開日 平成5年(1993)7月23日

(51) Int.Cl.⁵

G 0 6 F 15/40

識別記号

庁内整理番号

5 1 0 M 7060-5L

F I

技術表示箇所

審査請求 未請求 請求項の数1(全 9 頁)

(21) 出願番号 特願平3-345927

(22) 出願日 平成3年(1991)12月27日

(71) 出願人 000003078

株式会社東芝

神奈川県川崎市幸区堀川町72番地

(71) 出願人 000221052

東芝コンピュータエンジニアリング株式会社

東京都青梅市新町1381番地1

(72) 発明者 岩井 勇

神奈川県川崎市幸区柳町70番地 株式会社
東芝柳町工場内

(72) 発明者 中本 幸夫

東京都青梅市新町1381番地1 東芝コンピ
ュータエンジニアリング株式会社内

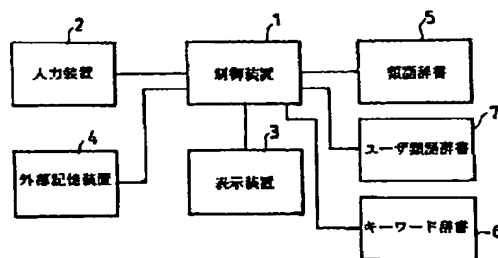
(74) 代理人 弁理士 三澤 正義

(54) 【発明の名称】 文書検索装置

(57) 【要約】

【目的】 ユーザの検索意図に合致した文書の検索を効率良く行える文書検索装置を提供する。

【構成】 ユーザが検索文を入力装置2から入力する。制御装置1は検索キーワードを切り出し、この検索キーワードを含む文書を検索する。該当する文書があった場合は、制御装置1は表示装置3にキーワードの類語一覧表を表示する。ユーザが入力装置2により表示された類語一覧表から類語を選択する。制御装置1は切り出したキーワードと選択された類語とを互いに関係付けてユーザ類語辞書部7のユーザ類語辞書に登録する。ユーザがユーザ類語辞書に登録されたキーワード又は類語を入力すると、制御装置1は互いに関係付けられたキーワード及び類語を検索キーワードとして文書の検索を行う。



【特許請求の範囲】

【請求項1】 複数の検索対象文書から検索キーワードを含む文書の検索を行う文書検索装置において、入力されたキーワードの類語一覧表を類語辞書を参照して表示する表示手段と、ユーザ類語辞書を登録するためのユーザ類語辞書部と、入力されたキーワードと前記類語一覧表から選択された類語とを互いに関係付けて前記ユーザ類語辞書部のユーザ類語辞書に登録すると共に、このユーザ類語辞書に登録され互いに関係付けられたキーワード又は類語のいずれかが入力された場合にそのキーワード及び類語を検索キーワードとして前記文書の検索を行う検索手段とを有することを特徴とする文書検索装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は、検索キーワードを含む文書の検索を行う文書検索装置に関する。

【0002】

【従来の技術】 ユーザがキーワードを入力してそのキーワードを含む文書を抽出する従来の文書検索装置は、ユーザが「検索装置の特色」のような自然文（以後「検索文」と呼ぶ）を入力すると、単語辞書によるマッチング処理または形態素解析処理により検索文から検索キーワードを切り出して、切り出した検索キーワードを含む文書を抽出するものである。

【0003】 しかし、検索キーワードとまったく同じ表記でなければ検索該当文書として抽出することはできない。従って、例えば「コンピュータ」と「計算機」のように類似の単語でも検索キーワードが「コンピュータ」で検索対象文書中が「計算機」の場合はその文書を抽出することができない。そこで、ユーザはこのような抽出漏れをなくすために、辞書引きなどして検索キーワードを追加していた。

【0004】

【発明が解決しようとする課題】 このように、従来の方式では、抽出漏れをなくすために検索キーワードの類語までを含む文書を抽出した場合には、ユーザの検索意図とはかけ離れたものを大量に与えてしまうという問題があった。また、ユーザが辞書によって的確な類語を選択して入力する方法も検索効率を下げ、ユーザの負担となる。

【0005】 また、ユーザが思考または辞書引きによって類語を選択し、検索キーワードをキーボードなどの入力手段を用いて入力し、その入力もand、orを指定しながら行うような場合は、このような作業はユーザにとって負担が大きいという問題があった。

【0006】 また、類語を計算機が自動処理し、類語をユーザ入力した検索キーワードにorとして追加し検索する場合や類語を一律に検索キーワードとする場合は、ユーザの検索意図とかけ離れたものまで抽出されることになる。例えば、「形式」の類語のなかには「形・作り

・外形・格好・形状・形質・異形・変形・形態」があり、個々に意味が異なる。ユーザが「何の形式」を求めているかによって類語が決まる。これらの語は確かに「形式」の類語であるがすべての語を検索キーワードとしてしまうのは、ユーザの検索意図とはかけ離れることになる。

【0007】 本発明は、上記事情に鑑みてなされたもので、ユーザの検索意図に合致した文書の検索を効率良く行える文書検索装置を提供することを目的とする。

【0008】

【課題を解決するための手段】 上記目的を達成するために本発明は、複数の検索対象文書から検索キーワードを含む文書の検索を行う文書検索装置において、入力されたキーワードの類語一覧表を類語辞書を参照して表示する表示手段と、ユーザ類語辞書を登録するためのユーザ類語辞書部と、入力されたキーワードと前記類語一覧表から選択された類語とを互いに関係付けて前記ユーザ類語辞書部のユーザ類語辞書に登録すると共に、このユーザ類語辞書に登録され互いに関係付けられたキーワード又は類語のいずれかが入力された場合にそのキーワード及び類語を検索キーワードとして前記文書の検索を行う検索手段とを有することを特徴とするものである。

【0009】

【作用】 ユーザがキーワードを入力すると、表示手段は入力されたキーワードの類語一覧表を表示する。ここで、ユーザが表示手段により表示された類語一覧表から類語を選択すると、検索手段はその入力されたキーワードと選択された類語とを互いに関係付けてユーザ類語辞書部のユーザ類語辞書に登録する。ユーザがユーザ類語辞書に登録されたキーワード又は類語を入力すると、検索手段は互いに関係付けられたキーワード及び類語を検索キーワードとして文書の検索を行う。

【0010】

【実施例】 図1は、本発明の文書検索装置の一実施例の概略構成を示すブロック図である。

【0011】 本装置は、CPU、メモリ等から構成されている検索手段としての制御装置1と、ユーザがキーボードやマウスなどによって検索文を入力したり検索操作をする入力装置2と、入力装置2から入力された検索文や検索操作の指示、検索結果及び類語選択画面を表示する表示手段としての表示装置3と、予めキーワードの類語辞書を格納している類語辞書部5と、ユーザ類語辞書を記憶するユーザ類語辞書部7と、キーワード辞書を格納するキーワード辞書部6と、検索対象文書などを格納する外部記憶装置4とから概略構成されている。

【0012】 前記制御装置1は、図2に示すように、初期化部201、データ読み込み部202、データ保存部203、キーワード切り出し部204、マッチング部205、類語選択画面表示部207、類語選択部208、類語反転表示部209、ユーザ類語辞書作成部210及び

び検索文生成部211の各処理部並びに一時記憶バッファ212、作業バッファ213及びキーワード格納バッファ214の各バッファから構成されている。

【0013】初期化部201は、各バッファ212乃至214の初期化を行うものである。データ読み込み部202は、外部記憶装置4からの検索で該当文書として抽出された検索対象文書の原文、類語辞書部5に格納している類語辞書及びユーザ類語辞書部7に格納しているユーザ類語辞書のバッファ212乃至214への読み込みを行うものである。データ保存部203は、外部記憶装置4へのデータ保存を行うものである。キーワード切り出し部204は、入力装置2から入力された検索文からキーワード辞書部6に格納されているキーワード辞書を参照し、検索キーワードを切り出してキーワード格納バッファ214に格納するものである。マッチング部205は、検索キーワードが検索対象文書に含まれているかどうかを判断するものである。

【0014】類語選択画面表示部207は、図4に示すように、キーワード格納バッファ214に格納されている単語（「検索」、「装置」、「特色」）の類語を類語辞書部5から読み込み、図8に示すように、表示装置3に表示するものである。

【0015】類語選択部208は、図9に示すように、類語選択画面表示部207により表示された検索キーワードの類語をマウスなどの位置入力手段によるユーザの選択操作に基づいて選択するものである。選択は、表示されている類語の近傍をユーザが指示することによって行われ、【設定終了】アイコンの指定により終了する。同図は「システム」、「特徴」又は「特長」が選択された場合を示す。

【0016】類語反転表示部209は、図8、図9に示すように、類語選択部208で選択された類語及びユーザ類語辞書部7のユーザ類語辞書に登録されている類語（「システム」、「特徴」及び「特長」）に対し、反転表示などの文字修飾を施すものである。

【0017】ユーザ類語辞書作成部210は、類語選択部208で選ばれた単語をユーザが入力したキーワードの類語を1グループとして作成し、ユーザ類語辞書部7にファイル保存するものである。

【0018】検索文生成部211は、類語選択部208で選ばれた類語及びユーザ類語辞書部7に保存されている類語と、ユーザが入力した検索文中の検索キーワードとの論理式を生成するものである。この論理式は、元の検索キーワードと選択された類語とを論理和としたものである。

【0019】前記表示装置3の初期画面は、図7に示すように、コマンドである【検索】、【類語表示】、【終了】を表示するようになっている。また、表示装置3の画面には、入力装置2から検索文が入力された場合は、その検索文をも表示するようになっている。同図の場合

では、「検索装置の特長」が入力された検索文として表示している。

【0020】前記類語辞書部5に予め記憶されている類語辞書は、図3に示すように、類語がグループ化されており各グループごとに番号が付られているものである。例えば同図では、「特徴」、「特色」、「特質」、「属性」、「個性」、「特長」及び「特点」からなるグループには20番が付られ、「電車」、「都電」、「市電」及び「地下鉄」からなるグループには31番が付られ、「画像」、「イメージ」及び「図」からなるグループには32番が付られている。ユーザ類語辞書部7に記憶されるユーザ類語辞書は、図6に示すように、ユーザが類語選択部208により類語辞書部5の類語辞書から選ばれた語がグループごとに番号が付られて格納されており、この番号は類語辞書部5の類語辞書の各グループごとに付られている番号に対応する。同図では、「装置」、「システム」、「特色」、「特徴」及び「特長」がユーザ類語辞書として登録されている例を示している。「装置」、「システム」からなるグループには15番が付られ、「特色」、「特徴」、「特長」からなるグループには20番が付られている。

【0021】次に、上記構成の本実施例の文書検索装置の作用を図10及び図11に示すフローチャートに従い説明する。

【0022】まず、各バッファを初期化する（S301）。【終了】ボタンがクリックされたかどうか判断して（S302）、クリックされれば終了する（S318）。ユーザがキーボードなどの入力装置2から検索文を入力する（S303）。例えば、「検索装置の特長」を検索文として入力する。次に【類語表示】アイコンがクリックされたかどうか判断する（S304）。このステップS304でクリックされていなければ、【検索】アイコンがクリックされたかどうかを判断する（S305）。前記ステップS304でキーワード切り出し部204が起動し、キーワード辞書部6に格納されているキーワード辞書を参照し、ユーザが入力した検索文から検索キーワードを切り出す（S311）。例えば、「検索装置の特徴」から「検索・装置・特徴」を切り出す。切り出したキーワードはキーワード格納バッファ214に格納する（図4）。バッファ214に検索キーワードを格納すると類語処理サブルーチン呼び（S312）、前記ステップS302に戻る。前記ステップS305でクリックされていなければ、前記ステップS302に戻る。前記ステップS305で【検索】アイコンがクリックされると、キーワード切り出し部204が起動し、キーワード辞書部6に格納されているキーワード辞書を参照し、ユーザが入力した検索文から検索キーワードを切り出す（S306）。そして、マッチング部205により、外部記憶装置4に格納されている検索対象文書から検索キーワードを含む文書を抽出する（S307）。そ

の結果を表示装置3に出力する(S308)。そこで該当文書があるかどうか判断する(S309)。このステップS309で該当文書が存在する場合は、前記ステップS302に戻る。前記ステップS309で該当文書がない場合は、類語サブルーチンと呼び(S310)、前記ステップS302に戻る。

【0023】類語サブルーチンは、検索キーワードが格納されているキーワード格納バッファ214を参照する(S311)。類語辞書(図3)を一時的記憶バッファ212に読み込み(S313)、検索キーワードの類語をサーチする(S313)。例えば、「特色」の検索キーワードから「特徴・特色・特質・属性・個性・特長・特点」がサーチされる。そして、類語選択画面表示部207が起動して、前記ステップS313でサーチした類語を表示する(S314、図8)。続いて類語反転表示部209が起動し、ユーザが入力したキーワードに対して先頭表示及び反転表示などを施し(図8)、キーワード格納バッファ212に格納されている単語に対しても反転表示などを施す。次に、ユーザ類語辞書部7に格納されているユーザ類語辞書(図6)にユーザ類語グループが登録されているか判断する(S315)。このステップS315でユーザ類語グループが登録されていないと、類語選択部208が起動してユーザはマウスなどの位置入力手段に類語を選択することができる。選択は表示されている類語の近傍を指定することによって行う。指定すると反転表示などの文字修飾などを施す(S317)。前記ステップS315でユーザ類語グループが登録されていると、ユーザが入力したキーワードを含むユーザ類語グループの単語も類語選択画面で反転表示などの文字修飾を施し(S316)、次のステップS317へ進む。【設定終了】アイコンが指定されるまでこのステップS317を繰り返す(S318)。【設定終了】アイコンが指定されると検索文生成部211が起動する。キーワード格納バッファ214に選択された語(反転表示などの文字修飾などが施されている語)を追加登録する(図5)。そして、キーワード格納バッファ214に格納されている語をユーザが入力した検索キーワードの論理和として検索論理式を生成し(S319)、キーワード格納バッファ214の内容を各キーワードの類語を1グループとしてユーザ類語辞書に格納する(S320)。これで、類語処理のサブルーチンを終了する。

【0024】このような上記実施例の文書検索装置によれば、以下の効果を奏する。

【0025】ユーザが入力した検索キーワードで検索対象文書中に該当文書が存在しない場合、自動的に検索キーワードの類語が表示され、ユーザが簡単操作で選択することによって検索キーワードを追加して検索すること

ができる。また、今までユーザが選択した類語をも記憶することにより、自動的に検索キーワードの類語を含めて検索処理する。この検索により、ユーザが類語を推測する必要なく、また、不要な語を検索キーワードとすることがない。

【0026】このように、検索キーワードの類語を表示することにより、ユーザの作業負担減と検索効率の向上が図れる。

【0027】なお、本発明は上記実施例に限定されず、その要旨を変更しない範囲内で種々に変形実施可能である。例えば、入力された検索文から切り出した検索キーワードを用いて文書の検索を行う場合について説明したが、本発明は、キーワードのみを入力した場合も同様に適用できる。

【0028】

【発明の効果】以上詳述したように本発明は、入力したキーワードの類語を容易に選択でき、そのキーワード及び類語を検索キーワードとして文書の検索を行うので、ユーザの検索意図に合致した文書の検索を効率良く行える文書検索装置を提供することができる。

【図面の簡単な説明】

【図1】本発明の文書検索装置の一実施例の概略構成を示すブロック図である。

【図2】制御装置の概略構成を示すブロック図である。

【図3】類語辞書の書式例を示す図である。

【図4】キーワード格納バッファの内容を示す図である。

【図5】キーワード格納バッファの内容を示す図である。

【図6】ユーザ類語辞書の一例を示す図である。

【図7】表示装置におけるユーザが入力した検索文表示例を示す図である。

【図8】表示装置における類語一覧表の表示例を示す図である。

【図9】表示装置における類語指定例を示す図である。

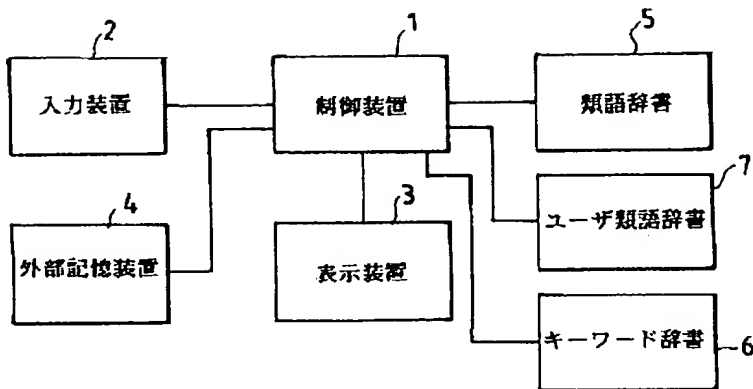
【図10】検索処理の流れを示すフローチャートである。

【図11】検索処理の流れを示すフローチャートである。

【符号の説明】

- 1 制御装置(検索手段)
- 2 入力装置
- 3 表示装置(表示手段)
- 4 外部記憶装置
- 5 類語辞書部
- 6 キーワード辞書部
- 7 ユーザ類語辞書部

【図1】



【図2】

初期化部	201
データ読み込み部	202
データ保存部	203
キーワード切り出し部	204
マッチング部	205
類語選択画面表示部	207
類語選択部	208
類語反転表示部	209
ユーザ類語辞書作成部	210
検索文生成部	211
一時記憶バッファ	212
作業バッファ	213
キーワード格納バッファ	214

【図4】

検索	
装置	
特色	

【図5】

検索	
装置	システム
特色	特徴 特長

【図7】

検索
類語表示
終了

検索文入力：検索装置の特色

(6)

特開平5-181912

【図3】

	:
2 0	特徴 特色 特質 属性 個性 特長 特点
	:
3 1	電車 都電 市電 地下鉄
3 2	画像 イメージ 図
	:

【図6】

1 5	装置	システム			
2 0	特色	特徴	特長		
2 2					
:					

【図8】

＜類語一覧表＞

☒ 体系 サーチ

☒ 装置 仕組み 構造 構図 方式 システム

☒ 特色 特徴 特質 属性 個性 特長 特点

設定終了

(7)

特開平5-181912

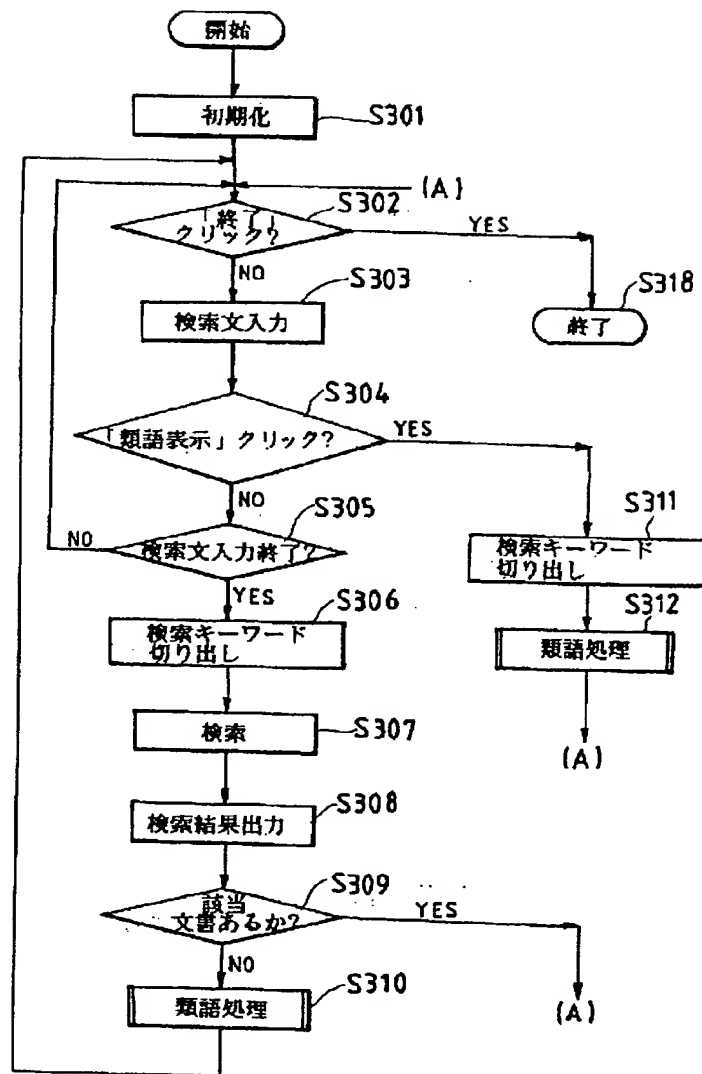
【図9】

〈類語一覧表〉

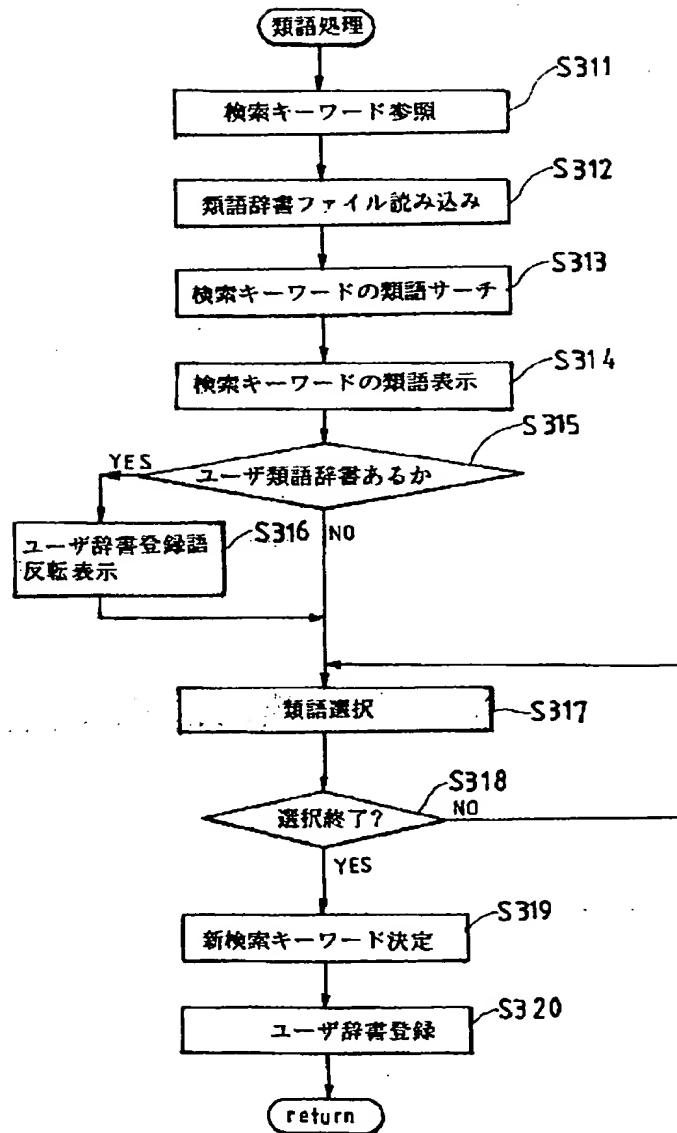
<input checked="" type="checkbox"/> 検査	サーチ	-----
<input checked="" type="checkbox"/> 装置	仕組み	構造 構図 方式 <input checked="" type="checkbox"/> システム
<input checked="" type="checkbox"/> 特色	<input checked="" type="checkbox"/> 特徴	特質 属性 個性 <input checked="" type="checkbox"/> 特長 特点

設定終了

【図10】



【図11】



This Page Blank (uspio)